Recitation 9. May 11

Focus: probability (discrete and continuous), random variables, principal component analysis (PCA)

A **random variable** is a quantity X that takes values in $\mathbb{R}$. It can be either:

- **discrete**: X takes only countably many possible values x_i each with probability p_i
- **continuous**: X is associated to a probability distribution $p(x)$ (where $p : \mathbb{R} \to \mathbb{R}$ is a function).

The **mean** (sometimes called "expected value") $E[X]$ of X is the quantity:

- $\boxed{\sum_i x_i p_i}$ if X is discrete
- $\boxed{\int_{-\infty}^{\infty} xp(x)\ dx}$ if X is continuous

The mean is linear: if X, Y are random variables and $a, b \in \mathbb{R}$, then $E[aX + bY] = aE[X] + bE[Y]$.

Given two random variables X, Y, their **covariance** $\Sigma_{XY} = E[(X - E[X])(Y - E[Y])]$ is:

- $\boxed{\sum_{ij} p_{ij}(x_i - \mu)(y_j - \nu)}$ if X is discrete
- $\boxed{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x - \mu)(y - \nu)p(x, y)\ dxdy}$ if X is continuous

The covariance of X with itself is called the **variance** Σ_{XX}.

Given n random variables $X_1, \dots, X_n$, we may assemble them into a vector $\boldsymbol{X} = \begin{bmatrix} X_1 \\ \vdots \\ X_n \end{bmatrix}$, called a **random vector**.

The **covariance matrix** of these random variables $X_1, \dots, X_n$ is the matrix

$$K = \begin{bmatrix} \Sigma_{X_1X_1} & \cdots & \Sigma_{X_1X_n} \\ \vdots & \ddots & \vdots \\ \Sigma_{X_nX_1} & \cdots & \Sigma_{X_nX_n} \end{bmatrix} = E[(\boldsymbol{X} - \boldsymbol{\mu})(\boldsymbol{X} - \boldsymbol{\mu})^T], \qquad \text{where} \quad \boldsymbol{\mu} = E[\boldsymbol{X}] = \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_n \end{bmatrix} = \begin{bmatrix} E[X_1] \\ \vdots \\ E[X_n] \end{bmatrix}$$

K is always positive semidefinite. It is positive definite unless a linear combination of $X_1, \dots, X_n$ is constant.

Principal component analysis (PCA) involves diagonalizing the covariance matrix:

$$\boxed{K = QDQ^T}$$

where Q is orthogonal and D is diagonal. This means that the random vector $\boldsymbol{Y} = Q^T\boldsymbol{X}$ has diagonal covariance matrix D, i.e. its entries are uncorrelated random variables (i.e. have covariance 0). In other words:

$$\boldsymbol{Y} = \begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} q_{11} & \cdots & q_{n1} \\ \vdots & \ddots & \vdots \\ q_{1n} & \cdots & q_{nn} \end{bmatrix} \begin{bmatrix} X_1 \\ \vdots \\ X_n \end{bmatrix} \qquad \Rightarrow \qquad \Big\{ Y_i = q_{1i}X_1 + \cdots + q_{ni}X_n \Big\}_{i \in \{1,\dots,n\}}$$

are linear combinations of $X_1, \dots, X_n$ that are (by construction) uncorrelated. The individual variances of the random variables $Y_1, \dots, Y_n$ are the diagonal entries of the diagonal matrix D.

1. Sample from the numbers 1 to 1000 with equal probabilities 1/1000, and look at the last digit of the sample, squared. This square can end with $X = 0, 1, 4, 5, 6$, or 9. What are the probabilities p_0, p_1, p_4, p_5, p_6 and p_9 that each of these digits occurs among the sample? Compute the mean and variance of X.

Solution: If $n = 10k$, then the last digit of n^2 will be 0. If $n = 10k + 1$ or $n = 10k + 9$, then the last digit of n^2 will be 1. If $n = 10k + 2$ or $n = 10k + 8$, then the last digit of n^2 will be 4. If $n = 10k + 3$ or $n = 10k + 7$, then the last digit of n^2 will be 9. If $n = 10k + 4$ or $n = 10k + 6$, then the last digit of n^2 will be 6. If $n = 10k + 5$, then the last digit of n^2 will be 5. Thus,

$$p_0 = \frac{1}{10} \qquad p_1 = \frac{1}{5} \qquad p_4 = \frac{1}{5} \qquad p_5 = \frac{1}{10} \qquad p_6 = \frac{1}{5} \qquad p_9 = \frac{1}{5}$$

We therefore see that the mean is

$$E[X] = 0 \cdot \frac{1}{10} + 1 \cdot \frac{1}{5} + 4 \cdot \frac{1}{5} + 5 \cdot \frac{1}{10} + 6 \cdot \frac{1}{5} + 9 \cdot \frac{1}{5} = \frac{9}{2},$$

and the variance

$$E\left[\left(X - \frac{9}{2}\right)^2\right] = \left(0 - \frac{9}{2}\right)^2 \frac{1}{10} + \left(1 - \frac{9}{2}\right)^2 \frac{1}{5} + \left(4 - \frac{9}{2}\right)^2 \frac{1}{5} + \left(5 - \frac{9}{2}\right)^2 \frac{1}{10} + \left(6 - \frac{9}{2}\right)^2 \frac{1}{5} + \left(9 - \frac{9}{2}\right)^2 \frac{1}{5} = \frac{181}{20}$$

2. Let A, H, and W denote random variables corresponding to the age, height, and weight of dogs at a local shelter, respectively. Suppose the random vector $\begin{bmatrix} A \\ H \\ W \end{bmatrix}$ takes two values, $\begin{bmatrix} 7 \\ 20 \\ 132 \end{bmatrix}$ and $\begin{bmatrix} 4 \\ 24 \\ 120 \end{bmatrix}$ with probabilities p and $1 - p$ respectively. Compute the covariance matrix of A, H, and W.

Solution: The mean of the random vector (i.e. the vector of means) is:

$$\boldsymbol{\mu} = \begin{bmatrix} \mu_A \\ \mu_H \\ \mu_W \end{bmatrix} = p \begin{bmatrix} 7 \\ 20 \\ 132 \end{bmatrix} + (1 - p) \begin{bmatrix} 4 \\ 24 \\ 120 \end{bmatrix} = \begin{bmatrix} 3p + 4 \\ 24 - 4p \\ 12p + 120 \end{bmatrix}$$

Then the covariances are given by:

$$\begin{aligned}
\Sigma_{AA} &= E[(A - \mu_A)^2] = p(7 - (3p + 4))^2 + (1 - p)(4 - (3p + 4))^2 = 9p(1 - p) \\
\Sigma_{HH} &= E[(H - \mu_H)^2] = p(20 - (24 - 4p))^2 + (1 - p)(24 - (24 - 4p))^2 = 16p(1 - p) \\
\Sigma_{WW} &= E[(W - \mu_W)^2] = p(132 - (12p + 120))^2 + (1 - p)(120 - (12p + 120))^2 = 144p(1 - p) \\
\Sigma_{AH} &= E[(A - \mu_A)(H - \mu_H)] = \\
&= p(7 - (3p + 4))(20 - (24 - 4p)) + (1 - p)(4 - (3p + 4))(24 - (24 - 4p)) = -12p(1 - p) \\
\Sigma_{AW} &= E[(A - \mu_A)(W - \mu_W)] = \\
&= p(7 - (3p + 4))(132 - (12p + 120)) + (1 - p)(4 - (3p + 4))(120 - (12p + 120)) = 36p(1 - p) \\
\Sigma_{HW} &= E[(H - \mu_H)(W - \mu_W)] = \\
&= p(20 - (24 - 4p))(132 - (12p + 120)) + (1 - p)(24 - (24 - 4p))(120 - (12p + 120)) = -48p(1 - p)
\end{aligned}$$

and so the covariance matrix is:

$$K = p(1 - p) \begin{bmatrix} 9 & -12 & 36 \\ -12 & 16 & -48 \\ 36 & -48 & 144 \end{bmatrix}$$

3. Suppose now that the random variables A, H, W from above instead have the covariance matrix

$$K = \begin{bmatrix} 3 & -1 & 2 \\ -1 & 3 & -2 \\ 2 & -2 & 6 \end{bmatrix}.$$

Find three linear combinations of A, H, W which are pairwise uncorrelated random variables. What is the variance of each?

Solution: We begin by diagonalizing K. Its characteristic polynomial is:

$$p_K(\lambda) = (3-\lambda)((3-\lambda)(6-\lambda)-4) + ((-1)(6-\lambda)+4) + 2(2-2(3-\lambda)) = (2-\lambda)^2(8-\lambda),$$

so the eigenvalues of K are 2 (with multiplicity 2) and 8. We now find a basis of eigenvectors. Since:

$$K - 8I = \begin{bmatrix} -5 & -1 & 2 \\ -1 & -5 & -2 \\ 2 & -2 & -2 \end{bmatrix}$$

from which we deduce that $\begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix}$ spans the null space of $K - 8I$. Thus, $\frac{1}{\sqrt{6}} \begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix}$ is an eigenvector (of norm 1) of K corresponding to eigenvalue 8. Similarly, we have that:

$$K - 2I = \begin{bmatrix} 1 & -1 & 2 \\ -1 & 1 & -2 \\ 2 & -2 & 4 \end{bmatrix}$$

from which we deduce that $\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$ and $\begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix}$ span the null space of $K - 2I$; moreover, these vectors are orthogonal (in this case, it was fairly easy to find a pair of orthogonal vectors spanning the null space by inspection, but in general you can always row reduce to find a basis for the null space and then apply Gram-Schmidt).

Thus, we have that $\frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$ and $\frac{1}{\sqrt{3}} \begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix}$ form an orthonormal basis for the eigenspace for eigenvalue 2. We therefore have:

$$\begin{aligned} K &= \begin{bmatrix} \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{3}} \\ -\frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}} \\ \frac{2}{\sqrt{6}} & 0 & \frac{1}{\sqrt{3}} \end{bmatrix} \begin{bmatrix} 8 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{3}} \\ -\frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}} \\ \frac{2}{\sqrt{6}} & 0 & \frac{1}{\sqrt{3}} \end{bmatrix}^T \\ &= \begin{bmatrix} \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{3}} \\ -\frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}} \\ \frac{2}{\sqrt{6}} & 0 & \frac{1}{\sqrt{3}} \end{bmatrix} \begin{bmatrix} 8 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{6}} & \frac{2}{\sqrt{6}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ -\frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \end{bmatrix}. \end{aligned}$$

This means that the random vector

$$\begin{bmatrix} \frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{6}} & \frac{2}{\sqrt{6}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ -\frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \end{bmatrix} \begin{bmatrix} A \\ H \\ W \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{6}}A - \frac{1}{\sqrt{6}}H + \frac{2}{\sqrt{6}}W \\ \frac{1}{\sqrt{2}}A + \frac{1}{\sqrt{2}}H \\ -\frac{1}{\sqrt{3}}A + \frac{1}{\sqrt{3}}H + \frac{1}{\sqrt{3}}W \end{bmatrix}$$

consists of random variables which are pairwise uncorrelated. Their variances are, respectively, 8, 2 and 2.

This process is known as principal component analysis. Note that because the covariance matrix in #2 has rank 2, it has 0 as an eigenvalue. Therefore, by a similar analysis we find that there must be a linear combination in that case of A, H, W which has variance 0, i.e. it is a constant.

4. Let X be a random variable, with mean μ and variance σ^2. Compute $E[X^2]$ in terms of μ and σ.

Solution: We have:

$$\sigma^2 = \Sigma_{XX} = E[(X-\mu)^2] = E[X^2 - 2\mu X + \mu^2] = E[X^2] - 2\mu E[X] + \mu^2 E[1] = E[X^2] - 2\mu^2 + \mu^2 = E[X^2] - \mu^2,$$

By adding μ^2 to both sides of the equation above, we get $E[X^2] = \sigma^2 + \mu^2$.